

VIDEO CONFERENCE AND VIDEO TELEPHONE SYSTEM,
TRANSMISSION APPARATUS, RECEPTION APPARATUS,
IMAGE COMMUNICATION SYSTEM, COMMUNICATION APPARATUS,
COMMUNICATION METHOD, RECORDING MEDIUM, AND PROGRAM

5

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to a video
conference and video telephone system which performs
10 multimedia communication based on packets, an image
communication system, a communication apparatus, a
communication method, a recording medium, and a
program.

Related Background Art

15 In a conventional video conference and video
telephone system, communication is performed mainly by
using an ISDN (Integrated Services Digital Network)
line on the basis of H.320 Standard of ITU-T
(International Telecommunication Union
20 Telecommunication standardization sector)
recommendation. In this system, the ISDN line must be
installed, and a toll of the ISDN line is expensive due
to its meter rate system. Thus, this system does not
spread a little, and usage thereof has been limited to
25 specific usage such as common usage within a conference
room of a company, or the like.

Against this, recently, a new standard for the

09870910-0314
FILED OCT 15 1990

video conference system called H.323 Standard of ITU-T recommendation in which a LAN (local area network) is used appeared, it came to be able to easily achieve the video conference by the LAN in the company. In this case, each user uses an H.323 video conference system corresponding to the LAN, whereby data communication can be performed without connection fees in the same LAN. Namely, only when the data communication is performed to the existing ISDN-based video conference system, such the data communication is performed through a common gateway, and thus the toll of the ISDN line is charged according to the meter rate system.

However, if there is a connection through the Internet and the other party also introduces the H.323 video conference system, the above gateway is unnecessary.

Further, since a faster LAN is advanced and thus a LAN based on 100Base-T of transfer rate 100Mbps class is also spreading, a connection of transfer rate 1Mbps class has been achieved in a local video conference connection, whereby image quality of such a video conference system is remarkably improved as compared with that of a conventional video conference system of 2B 128Kbps based on the ISDN.

Further, since spreading of the faster Internet started, connection speed between the LAN's is rapidly improved. Thus, when the video conference between the

09870910-053101

[illegible]

10 Since the line toll increases in proportion to the
number of conference participants in the conventional
ISDN-based H.320 system, this system has an extremely
luxurious function when thinking about costs of the
communication lines. Further, since the band of the
15 line is narrow, communication quality is unexcellent.

20 Further, when it pays attention to the point of
audio (or voice), the ISDN-based H.320 system is a
standard only of monaural. For this reason, when it is
intended to achieve stereo, audio data (or voice data)
erodes the band of video data (or image data) in case
25 of the basic 2B connection, whereby image quality is
deteriorated. On the other hand, in the LAN-based
H.323 system, particularly in the same LAN, since a

data transfer rate is high (10Mbps, 100Mbps, etc.), on the data transfer any serious problem is not caused even if the band increases because the audio data is made stereo.

5 Thus, if it is intended to make the audio data stereo and achieve the group conference, later-described problems are caused in the specification described in the latest H.323 Standard Book (TTC (Telecommunication Technology Committee) Standard JT-
10 H.323 Ver. 2.1).

A group telephone and communication system includes two systems, i.e., a centralized multipoint connection system, and a non-centralized multipoint connection system.

15 First, in the group conference system, the non-centralized multipoint connection system which can be most simply achieved will be explained hereinafter by way of example. In H.323 Standard, since video and audio are transmitted/received respectively on
20 independent packets, the explanation of the video will be omitted here.

Fig. 5 shows a configuration of the non-centralized multipoint connection. In case of the non-centralized multipoint connection, for example, a case
25 where there are three participants A, B and C is thought. In Fig. 5, a termination point which generates an information stream of a terminal (i.e.,

00870910-053101

the participant) A is shown as an end point A 501.

Similarly, a termination point which generates an information stream of a terminal (i.e., the participant) B is shown as an end point B 502, and a termination point which generates an information stream of a terminal (i.e., the participant) C is shown as an end point C 503. When the multipoint connection is performed, a multipoint controller (MC) which performs multipoint control is necessary. The function of this MC may be achieved by a multipoint processor (MPU) or the terminal itself participating in the conference. In Fig. 5, although a MC 504 is independently shown for intelligibility, it is assumed that the MC is actually included in the terminal (the end point).

The terminal A notifies beforehand each participant of holding of the group conference by means of, e.g., an electronic mail or the like. The MC 504 existing in the terminal A performs setting to convene the conference. Next, the end point A 501 performs call setting to the MC 504. After the call setting was performed, the end point A 501 performs capability exchange to other terminals according to H.245 Standard for a multimedia communication control protocol.

The end point B 502 and the end point C 503 being other participants also perform call setting to the MC 504 and perform capability exchange to other terminals according to H.245 Standard. The MC 504 gathers and

09870910.053404
T01550.01607860

composites all participants' capabilities, selects therefrom the common capability, e.g., audio according to G.711 Standard for an audio compression system as a selection communication mode (SCM) in this case,

5 transmits the selected SCM by using a communication mode command, describes this SCM in a communication mode table 520, and then transmits the SCM to the respective end points through 507, 508 and 509. It should be noted that this SCM is described in the

10 communication mode table 520 in the form of an entry 1.

The contents of the communication mode table 520 include SESSION ID (= 1) representing a session, SESSION DESCRIPTION (= audio) representing session contents, DATA TYPE (= G.711 monaural) representing a

15 data type, MEDIA CHANNEL (= MCA 1 505) representing a multicasting address for transmitting audio data, and MEDIA CONTROL CHANNEL (= MCA 2 506) representing a multicasting address for transmitting audio control data.

20 Then, each participant's terminal starts to transmit audio and thus starts multicasting. The end point A 501 transmits the audio data to the MCA 1 505 through 510, and transmits the audio control data to the MCA 2 506 through 513.

25 Similarly, the end point B 502 transmits the audio data to the MCA 1 through 511, and transmits the audio control data to the MCA 2 through 514. Further, the

09870910-053101

end point C 503 transmits the audio data to the MCA 1 through 512, and transmits the audio control data to the MCA 2 through 515.

For example, the end point A 501 receives
5 multicasting audio channels, executes an audio mixing function, and thus can provide a composited audio signal to the user.

As described above, a non-centralized multipoint conference is completed. If the terminal A being the
10 convener performs end setting, the conference ends. Of course, the participant can arbitrarily retire from the conference but can not end the conference. The above is the operation of the non-centralized multipoint conference using monaural audio.

15 On the other hand, in the centralized multipoint connection system, a multipoint control unit (MCU) or a terminal capable of achieving an MCU function is necessary. In this conference, each of all the terminals participating in the group telephone and
20 conference communicates with the MCU in a point-point manner. Each terminal transmits its control stream, audio stream, video stream and data stream to the MCU. The MCU performs various processes such as compositing and the like to the received data, and then transmits
25 the processed data to each terminal.

Further, in the centralized multipoint connection system, each participant's terminal multicasts audio

09870910-053101
T0750 07602860

data and video data to all of other participant's terminals. It is necessary for each terminal to composite the received audio stream and select one or plural video streams to be displayed.

5 Besides, there is a mixing multipoint connection system that the plural group telephone and conference systems are appropriately combined. In this system, the plural terminals participating in this conference in the centralized multipoint connection system and the
10 plural terminals participating in this conference in the non-centralized multipoint connection system together perform the group telephone and conference.

 In the video telephone and conference using H.323 Standard, since each of the audio stream and the video
15 stream is transmitted/received in the form of independent packet, only the audio will be explained hereinafter.

 Fig. 15 shows topology of the group telephone and conference according to the centralized multipoint
20 connection. In this centralized multipoint connection, as described above, an MCU 1601 is necessary. In this group telephone and conference, each of participant three terminals A 1602, B 1603 and C 1604 communicates with the MCU 1601 in a point-point manner.

25 Generally, the MCU has one multipoint controller (MC) and plural multipoint processors (MP's). The MCU 1601 in Fig. 15 has one MC and one MP managing the

09870910-053101

audio data.

In order to perform the group conference, the MC existing in the MCU performs setting to convene the conference. Each of the terminals A 1602, B 1603 and C 1604 participating in this conference first performs call setting to the MC, and then performs capability exchange to other terminals according to H.245 Standard. Thus, the MC gathers and composites all participants' capabilities, selects therefrom the common capability as a selection communication mode (SCM).

Then, each terminal transmits the audio data to the MCU by using the SCM determined as a result of the capability exchange.

The MP in the MCU performs a gathering process to the audio data received from the respective terminals. The MP composites the plural received audio data, performs a predetermined process thereto, and then multicasts the audio data converted to the SCM mode to the respective terminals.

When the MCU being the convener of the conference performs end setting, the conference ends. Of course, each participant's terminal can arbitrarily retire from the conference but can not end the conference.

On the other hand, if it is intended to perform the multipoint conference that the audio is made stereo, following problems are caused. Namely,

09870910-053104
TOTCS9" 07692960

5

10

15

20

25

control protocol) is newly added to the standard, the terminal A notifies the terminal B that the terminal A was changed to the monaural audio source, the stereo audio connection is disconnected, and the monaural audio connection is reset, then the audio can be made monaural to save the band. However, in this case, there is a drawback that the processing operation becomes complex.

It is rare that all the terminals participating in the group telephone and conference have the same processing capability. For example, when it pays attention to the number of audio channels, the terminals A and B are the terminals each having the stereo signal processing capability, and the terminal C is the terminal having the monaural signal processing capability. Thus, at this time, data 1605 transmitted from the terminal A 1602 to the MCU 1601 is the stereo audio composed of L audio data and R audio data, data 1606 transmitted from the terminal B 1603 to the MCU 1601 is the stereo audio composed of L audio data and R audio data, and data 1607 transmitted from the terminal C 1604 to the MCU 1601 is a monaural signal. Thus, in this group telephone and conference, the MCU 1601 multicasts audio data 1608 in which the signals obtained by making the audio signals of the terminals A and B monaural and the audio signal of the terminal C have been added to others.

09870910-053101
TOP SECRET 07607860

As described above, when the group telephone and conference is performed in the situation that the stereo and monaural terminals mixedly exist, even if the terminal (e.g., the terminal A or B) has the stereo
5 signal processing capability, this terminal can do nothing but receive the monaural signal.

SUMMARY OF THE INVENTION

09870910-053101
10 An object of the present invention is to solve all or at least one of the above problems.

Another object of the present invention is to achieve a video conference and video telephone system which solves the above problems and makes audio stereo.

15 Still another object of the present invention is to provide a system which deals with stereo audio as a whole irrespective of whether each terminal constituting this system deals with stereo audio or monaural audio, and thus efficiently uses lines.

20 Under the above objects, according to one aspect of the present invention, it is provided a video conference and video telephone system which includes transmission and reception apparatuses for performing communication of two audio signals of L and R channels, wherein

25 the transmission apparatus comprises

a transmission means for transmitting data obtained by addition of the two audio signals as first

audio data through a first communication channel, and transmitting data obtained by subtraction of the two audio signals as second audio data through a second communication channel, and

5 the reception apparatus comprises

a reception means for receiving the data obtained by the addition of the two audio signals as the first audio data and the data obtained by the subtraction of the two audio signals as the second
10 audio data, and

a restoring means for restoring the audio signal by performing an arithmetic operation on the basis of the audio data received by the reception means.

15 According to another aspect of the present invention, it is provided a transmission apparatus in a video conference and video telephone system which has a transmission means for transmitting packet data obtained by addition of two audio signals of L and R
20 channels through a first communication channel, and transmitting packet data obtained by subtraction of the two audio signals through a second communication channel.

According to still another aspect of the present
25 invention, it is provided a reception apparatus in a video conference and video telephone system which has a reception means for receiving packet data obtained by

09870910-053101
TOP SECRET

addition of two audio signals of L and R channels
and/or packet data obtained by subtraction of the two
audio signals, and a restoring means for restoring the
audio signal by performing an arithmetic operation on
5 the basis of the packet data received by the reception
means.

According to still another aspect of the present
invention, it is provided a communication apparatus
which has a transmission means for transmitting packet
10 data obtained by addition of two audio signals of L and
R channels through a first communication channel and
packet data obtained by subtraction of the two audio
signals through a second communication channel, a
reception means for receiving the packet data obtained
15 by the addition of the two audio signals of the L and R
channels and/or the packet data obtained by the
subtraction of the two audio signals, and a restoring
means for restoring the audio signal by performing an
arithmetic operation on the basis of the packet data
20 received by the reception means.

According to still another aspect of the present
invention, it is provided a communication method which
has a step of transmitting packet data obtained by
addition of two audio signals of L and R channels
25 through a first communication channel and packet data
obtained by subtraction of the two audio signals
through a second communication channel.

09870910-053104

According to still another aspect of the present invention, it is provided a communication method in a video conference and video telephone system which has a step (a) of receiving packet data obtained by addition of two audio signals of L and R channels and/or packet data obtained by subtraction of the two audio signals, and a step (b) of restoring the audio signal by performing an arithmetic operation on the basis of the packet data received in the step (a).

According to still another aspect of the present invention, it is provided a communication method which has a step (a) of transmitting packet data obtained by addition of two audio signals of L and R channels through a first communication channel and packet data obtained by subtraction of the two audio signals through a second communication channel, a step (b) of receiving the packet data obtained by the addition of the two audio signals of the L and R channels and/or the packet data obtained by the subtraction of the two audio signals, and a step (c) of restoring the audio signal by performing an arithmetic operation on the basis of the packet data received in the step (b).

According to still another aspect of the present invention, it is provided a computer-readable recording medium which records therein a program to cause a computer to execute a procedure of transmitting packet data obtained by addition of two audio signals of L and

R channels through a first communication channel and packet data obtained by subtraction of the two audio signals through a second communication channel.

According to still another aspect of the present invention, it is provided a computer-readable recording medium which records therein a program to cause a computer to execute a procedure (a) of receiving packet data obtained by addition of two audio signals of L and R channels and/or packet data obtained by subtraction of the two audio signals, and a procedure (b) of restoring the audio signal by performing an arithmetic operation on the basis of the packet data received in the procedure (a).

According to still another aspect of the present invention, it is provided a computer-readable recording medium which records therein a program to cause a computer to execute a procedure (a) of transmitting packet data obtained by addition of two audio signals of L and R channels through a first communication channel and packet data obtained by subtraction of the two audio signals through a second communication channel, a procedure (b) of receiving the packet data obtained by the addition of the two audio signals of the L and R channels and/or the packet data obtained by the subtraction of the two audio signals, and a procedure (c) of restoring the audio signal by performing an arithmetic operation on the basis of the

05870510 DE3101
FOTCSO DT60650

packet data received in the procedure (b).

According to the present invention, by performing the communication for the data obtained by the addition of the two audio signals of the L and R channels and the data obtained by the subtraction of the two audio signals, it is possible to deal with both the stereo audio and the monaural audio. In the multipoint conference in which the terminals each having the stereo audio processing capability and the terminals each having the monaural audio processing capability mixedly participate, it is possible between the terminals each having the stereo audio processing capability to restore the stereo audio without increasing a data quantity and wastefully increasing processing capabilities.

In the present invention, it is disclosed an image communication system which is composed of transmission and reception apparatuses performing communication of two audio signals of L and R channels, wherein the transmission apparatus comprises

a reception means for receiving, from an external apparatus, the two audio signals of the L and R channels and a monaural audio signal,

a transmission means for transmitting data obtained by addition of the received two audio signals and monaural audio signal as first audio data through a first communication channel, and transmitting data

09870910.053101
TOP SECRET

obtained by subtraction of the two audio signals as second audio data through a second communication channel, and

the reception apparatus comprises

5 a reception means for receiving the data obtained by the addition of the two audio signals and monaural audio signal as the first audio data and the data obtained by the subtraction of the two audio signals as the second audio data, and

10 a restoring means for restoring a stereo audio signal on the basis of the first and second audio data received by the reception means.

Further, in the present invention, it is disclosed a communication apparatus which performs communication
15 with plural external apparatuses, comprising:

 a reception means for receiving, from the external apparatus, two audio signals of L and R channels or a monaural audio signal;

 a generation means for generating first audio data
20 by addition of the received two audio signals and monaural audio signal and second audio data by subtraction of the two audio signals; and

 a transmission means for transmitting the first and second audio data.

25 Further, in addition to the above structure, it is disclosed a communication apparatus wherein the transmission means transmits the first audio data

09870910 053101

through a first communication channel and the second audio data through a second communication channel.

Further, in addition to the above structure, it is disclosed a communication apparatus wherein, when the
5 external apparatus at a transmission destination of the transmission means corresponds to stereo audio, the transmission means transmits the first and second audio data to the transmission destination, and when the
10 external apparatus at the transmission destination of the transmission means corresponds to monaural audio, the transmission means transmits the first audio data to the transmission destination without transmitting the second audio data.

Further, in addition to the above structure, it is
15 disclosed a communication apparatus which further comprises an image data communication means for transmitting and receiving image data.

Further, in the present invention, it is disclosed a communication method for an image communication
20 system which is composed of transmission and reception apparatuses performing communication of two audio signals of L and R channels, wherein

in the transmission apparatus, the method comprises
25 a reception step of receiving, from an external apparatus, the two audio signals of the L and R channels and a monaural audio signal, and

09870910-053101

09870910-053101

a transmission step of transmitting data
obtained by addition of the received two audio signals
and monaural audio signal as first audio data through a
first communication channel, and transmitting data
5 obtained by subtraction of the two audio signals as
second audio data through a second communication
channel, and

in the reception apparatus, the method further
comprises
10 a reception step of receiving the data
obtained by the addition of the two audio signals and
monaural audio signal as the first audio data and the
data obtained by the subtraction of the two audio
signals as the second audio data, and
15 a restoring step of restoring a stereo audio
signal on the basis of the first and second audio data
received in the reception step.

Further, it is disclosed a communication method
for a communication apparatus which performs
20 communication with plural external apparatuses,
comprising:

a reception step of receiving, from the external
apparatus, two audio signals of L and R channels or a
monaural audio signal;
25 a generation step of generating first audio data
by addition of the received two audio signals and
monaural audio signal and second audio data by

subtraction of the two audio signals; and

a transmission step of transmitting the first and second audio data.

Further, in addition to the above structure, it is disclosed a communication method wherein the transmission step transmits the first audio data through a first communication channel and the second audio data through a second communication channel.

Further, in addition to the above structure, it is disclosed a communication method wherein

when the external apparatus at a transmission destination in the transmission step corresponds to stereo audio, the transmission step transmits the first and second audio data to the transmission destination, and

when the external apparatus at the transmission destination in the transmission step corresponds to monaural audio, the transmission step transmits the first audio data to the transmission destination without transmitting the second audio data.

Further, in addition to the above structure, it is disclosed a communication method wherein an image data communication step of transmitting and receiving image data is provided.

Further, it is disclosed a program which causes a computer to achieve a communication method comprising:

a first generation step of generating packet data

09870910.053101

obtained by addition of two audio signals of L and R channels;

a second generation step of generating packet data obtained by subtraction of the two audio signals; and

5 a transmission step of transmitting the packet data generated in the first generation step through a first communication channel, and transmitting the packet data generated in the second generation step through a second communication channel.

10 Other objects and features of the present invention will be clarified through the following description in the specification and the attached drawings.

15 BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram showing a video conference and video telephone system according to the embodiment of the present invention;

20 Fig. 2 is a block diagram showing a stereo audio circuit;

Fig. 3 is a schematic diagram of the video conference and video telephone system according to the first embodiment;

25 Fig. 4 is a block diagram showing a process in an audio DSP (digital signal processor);

Fig. 5 is a schematic diagram of conventional non-centralized multipoint connection;

00870910-053104

Fig. 6 is a schematic diagram of non-centralized multipoint connection according to the first embodiment;

Fig. 7 is a diagram showing an example of a
5 capability table according to the first embodiment;

Fig. 8 is a diagram showing an example of a capability table of a terminal having a monaural audio processing capability;

Fig. 9 is a diagram showing an example of an RTCP
10 (real time control protocol) sender report packet which is transmitted by the system of the first embodiment;

Fig. 10 is a block diagram showing an audio process in an MCU (multipoint control unit) according to the second embodiment;

Fig. 11 is an internal block diagram showing a
15 stereo video telephone and conference terminal according to the second embodiment;

Fig. 12 is a block diagram showing an internal audio data process in the stereo video telephone and
20 conference terminal according to the second embodiment;

Fig. 13 is a block diagram showing an internal audio data process in a monaural video telephone and conference terminal;

Fig. 14 is a schematic diagram of group telephone
25 and conference which uses centralized multipoint connection according to the second embodiment;

Fig. 15 is a schematic diagram of group telephone

09870910-053101

and conference which uses conventional centralized multipoint connection;

Fig. 16 is a diagram showing a capability table of a stereo video telephone and conference terminal;

5 Fig. 17 is a diagram showing a main audio data packet which is multicast by an MCU;

Fig. 18 is a diagram showing a sub audio data packet which is multicast by the MCU;

10 Fig. 19 is a schematic diagram of group telephone and conference which uses the centralized multipoint connection according to the second embodiment; and

Fig. 20 is a block diagram showing an internal audio data process in the video telephone and conference terminal having an MCU function, according to the second embodiment.

15

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[First Embodiment]

20 The first embodiment of the present invention will be explained hereinafter. A video conference and video telephone system according to the present embodiment has a means which performs a following process in audio data communication.

25 A transmission side performs an arithmetic operation based on L and R audio signals to generate an $(L+R)/2$ signal and $(L-R)/2$ signal, and performs encoding.

09870910.053101

Then, on a first audio channel, the transmission side transmits, as standard monaural audio, audio data obtained by encoding the $(L+R)/2$ signal. On the other hand, on a second audio channel, the transmission side
5 transmits, as nonstandard data, audio data obtained by encoding the $(L-R)/2$ signal.

In the video conference and video telephone system on a reception side, a terminal which merely has only a monaural audio reception capability or a terminal which
10 wishes to dare to receive the transmitted data as monaural audio receives $(L+R)/2$ data being the monaural audio on the first audio channel, and decodes the received data to restore or reproduce the audio on the transmission side.

15 A terminal which wishes to receive the stereo audio receives $(L+R)/2$ data being the monaural audio and $(L-R)/2$ data being the nonstandard data on the second audio channel.

Then, data compositing is performed by using time
20 stamps of the $(L+R)/2$ and $(L-R)/2$ data, and the composited data is decoded. The $(L+R)/2$ and $(L-R)/2$ signals obtained by the decoding are subjected to addition and subtraction processes, whereby the audio on the L and R channels on the transmission side is
25 restored.

By the above means, in the multipoint conference in which the terminals each having the stereo audio

09870910.053404
T07650.07692860

processing capability and the terminals each having the monaural audio processing capability mixedly participate, it is possible between the terminals each having the stereo audio processing capability to
5 restore the stereo audio without increasing a data quantity and wastefully increasing processing capabilities.

Further, it is provided the function which controls connection/non-connection of the second audio
10 channel according to whether the audio input source is the monaural audio input source or the stereo audio input source. Further, notification of such the audio source change is described in a command of H.245
Standard or a capability table, or uses an SDES (source
15 description) of an RTCP (real time control protocol) packet. Thus, between the terminals each having the stereo transmission/reception capability, it is possible to control the connection/non-connection of the second audio channel according to the audio source
20 change between monaural and stereo, whereby the band can be efficiently used.

First, an example of hardware of the video conference and video telephone system according to the embodiment of the present invention will be explained
25 with reference to the attached drawings. Next, an operation in a case where the multipoint connection video conference is performed with use of the video

09870910 053404
T0T50 0T602860

conference and video telephone system of the above hardware will be explained. Fig. 1 is a block diagram showing the video conference and video telephone system according to the present embodiment, and Fig. 3 is a
5 schematic diagram of this video conference and video telephone system.

In Fig. 1, when power is supplied from a power supply 116 to the video conference and video telephone system, s system controller 105 reads a predetermined
10 program code for system operation from a flash ROM 107, loads the read program code to an SDRAM (synchronous dynamic random access memory) 108, and actually executes a program. By this program, each block constituting the system is reset and then set to a
15 predetermined initial state. After a video codec (coder-decoder) 103 was reset, the program code for the video codec 103 is read by the system controller 105 from a predetermined area of the flash ROM 107, and the read code is loaded to a not-shown SRAM (static random
20 access memory) in the video codec 103. Subsequently, a predetermined command is sent from the system controller 105 to the video codec 103 to start the loaded program. A similar operation is performed by the system controller 105 to an audio codec 104. After
25 such a series of initialization operations at the start time, the video conference and video telephone system can enter into an ordinary operation state.

09870910-053101

After the video conference and video telephone system entered into the ordinary operation state, this system performs a following operation. Namely, as to audio input, an analog video output image generated by a video camera 302 of a terminal 301 of Fig. 3 is supplied to a video decoder 101 (CAMERA IN). Since the video decoder 101 is a multi-input type, plural kinds of video cameras are selectable. In case of selecting one of plural input video signals, for example, a predetermined control signal is sent through a wireless unit 110 from the system controller 105 of Fig. 1 to the video decoder 101 on the basis of a selection signal from an operation switch provided on an operation unit 308 of Fig. 3.

An input video signal from a selected input source is digitized and sent to the video codec 103 by the video decoder 101. Then, in the video codec 103, the obtained digital video signal is subjected to a predetermined process, and an image data quantity is compressed according to a video compression algorithm based on, e.g., H.261 Standard recommended by ITU-T.

On the other hand, with respect to the audio input, for example, an audio signal which was sent from stereo microphones 303 and 304 (MIC IN), an external line (AUDIO LINE IN), a headset (HEADSET), a wireless telephone 309 through the wireless unit 110, or the like is supplied to an audio input selector 113

partially through a stereo circuit 114, and an arbitrary audio input is selected by the selector 113. The audio input selected by the audio input selector 113 is input to an audio AD/DA (analog-to-
5 digital/digital-to-analog) converter 112.

The selection of the audio input source is controlled according as a command is sent from the system controller 105 to a control latch circuit 115 on the basis of an user's working.

10 The audio signal digitized by the audio AD/DA converter 112 is supplied to the audio codec 104. In the audio codec 104, the obtained digital audio signal is subjected to an audio data compression process based on, e.g., G.711 Standard recommended by ITU-T.

15 When the video conference is performed over the LAN, the video and the audio are transmitted respectively as different packet data on the basis of H.323 Standard recommended by ITU-T, and they are synchronized with each other by using respective time
20 stamps.

Thus, the video signal compressed by the video codec 103 is sent to the system controller 105, subjected to predetermined fragmentation based on H.225.0 Standard recommended by ITU-T, and then
25 subjected to a predetermined process to create the packet data. On the other hand, the audio signal compressed by the audio codec 104 is similarly sent to

09370910-05101

the system controller 105, subjected to predetermined fragmentation based on H.225.0 Standard recommended by ITU-T, and then subjected to a predetermined process to create the packet data. Each of the video and audio packet data is transmitted from the system controller 105 to a LAN line through a LAN I/F (interface) 109, and the transmitted data packet is received by a video conference system at a transmission destination, whereby predetermined video and audio are reproduced on this system.

On the other hand, packet data fragmentated based on H.225.0 Standard for partner's video and audio are transmitted from an opposed video conference system and received by the system controller 105 through the LAN I/F 109. In the system controller 105, the fragmentated packet data are restructured respectively into video and audio compression data and then synchronized by using respective time stamps. The restructured compression video data is decoded and restored into the original video signal by the video codec 103.

On the other hand, the restructured audio signal is decoded and restored into the original audio signal by the audio codec 104.

The restored video signal is displayed on a monitor 305. The restored audio signal is converted into the analog audio signal by the audio AD/DA

09870910 053404
T0750 07604960

5

10

15

20

09870910.053101
a switch 204 is turned on, while if the headset is selected as the input source, a switch 205 is turned on. The switches 204 and 205 are controlled by the system controller 105 with use of the control latch
5 circuit 115.

Further, in the video conference and video telephone system, there are three audio output routes including the wireless unit (wireless telephone) 202, the headset (HEADSET) through the headset connector
10 203, and an audio line out (AUDIO LINE OUT). With respect to a signal to be supplied to the telephone or the headset which acts as a monaural output, in consideration of its band, stereo outputs from L (LOUT) and R (ROUT) channels of the audio AD/DA unit 201 are
15 added by an adder 210, band-limited by an LPF (low-pass filter) 211 of 3kHz, and then output to the telephone or the headset. Further, the stereo outputs from the audio AD/DA unit 201 are output respectively to L (LOUT) and R (ROUT) channels of a terminal (AUDIO LINE
20 OUT) capable of performing stereo outputs.

In a case where the system on the user's own side is selecting a VTR (video tape recorder) audio input, not only the audio on the partner's side (other station) being in the video conference and video
25 telephone communication but also the audio of a VTR must be added to the system audio output. For this reason, when the VTR is used as the audio input source,

a switch 212 is turned on, the VTR audio signal is added to the signal output from the audio AD/DA unit 201 by R- and L-channel adders 208 and 209, and the obtained signal is then output from the speaker or the
5 like as the audio output of the video conference system.

Fig. 4 is a block diagram showing a stereo audio signal process in a DSP (digital signal processor) which processes the audio signal within the system. In
10 order to transmit the stereo audio, following signal processes on blocks are performed.

An L-channel audio signal 401 and an R-channel audio signal 402 are input to an audio signal arithmetic block 403. In the audio signal arithmetic
15 block 403, size-adjusted arithmetic signals, i.e., an $(L+R)/2$ signal 404 and an $(L-R)/2$ signal 405, are obtained and output. The $(L+R)/2$ signal 404 is then encoded by a codec block 406, and encoded $(L+R)/2$ data 408 is output. This $(L+R)/2$ data can be managed as a
20 conventional monaural audio signal and is called a standard audio signal.

The $(L-R)/2$ signal 405 is encoded by a codec block 407, and encoded $(L-R)/2$ data 409 is output. The output $(L-R)/2$ data 409 can not be managed as the
25 conventional monaural audio signal (i.e., the standard audio signal) in this video conference system. Thus, the output $(L-R)/2$ data 409 is transmitted together

00070910.053101

with discrimination information as a nonstandard audio signal.

Next, in order to receive the above generated stereo audio data, following signal processes on blocks
5 are performed. Namely, the received audio data of the two channels have been synchronized with each other by the system controller 105, and the received audio data is thus decoded and subjected to an arithmetic operation as follows in the audio DSP.

10 The received monaural audio data, i.e., $(L+R)/2$ data 410, is decoded by a codec block 412, and a decoded $(L+R)/2$ audio signal 414 is output.

Further, the received nonstandard audio signal, i.e., $(L-R)/2$ signal 411, is decoded by a codec block
15 413, and a decoded $(L-R)/2$ audio signal 415 is output. The decoded $(L+R)/2$ audio signal 414 and the decoded $(L-R)/2$ audio signal 415 are input to an audio signal arithmetic block 416. In the audio signal arithmetic block 416, the input signals are subjected to addition
20 and subtraction processes, whereby an L-channel signal 417 and an R-channel signal 418 being the audio signals on the partner's side are restored.

Next, a multipoint conference which uses the video conference system according to the present embodiment
25 will be explained hereinafter. Fig. 6 shows non-centralized multipoint connection which uses the video conference system according to the present embodiment.

09870910-053101
T07250" 07607860

It is assumed that there are three terminals (parties) A, B and C in the non-centralized multipoint connection.

5 In Fig. 6, a point of the terminal A which generates and terminates its information stream is called an end point A 601. Similarly, points of the terminals B and C which generate and terminate their information streams are called end points B 602 and C 603, respectively. When the multipoint connection is performed, a multipoint controller (MC) 604 is necessary. However, a multipoint processor (MPU) or the terminal participating in the conference may have a function of the MC 604. In Fig. 6, although the MC 604 is independently shown for intelligibility, it is assumed that the MC 604 is actually included in the terminal A.

The terminal A notifies beforehand each participant of holding of the group conference by means of, e.g., an electronic mail or the like. The terminal A performs setting to convene the conference for the MC 604. Next, the end point A 601 performs call setting to the MC 604. After the call setting was performed, the end point A 601 performs capability exchange to other terminals according to H.245 Standard.

25 Fig. 7 shows an example of a capability table of the end point A 601 which is used in the capability exchange. In this case, it is assumed that the video

09870910.053101
F0T59"0T692860

conference system at the terminal A has a stereo audio processing capability. In Fig. 7, a description 701 indicates a data conference capability, an environment to be used, and the like, a description 702 indicates a capability for receiving audio G.711A-LAW compressed based on G.711A-LAW Standard being one of audio signal compression systems, and a description 703 indicates a capability for receiving audio G.711U-LAW. The capabilities indicated by the descriptions 702 and 703 aim at monaural audio of one channel. In this system, the $(L+R)/2$ audio data is transmitted in this channel.

A description 704 indicates nonstandard audio data. Here, the $(L-R)/2$ audio data encoded based on G711A-LAW Standard is managed.

A description 705 indicates nonstandard audio data. Here, the $(L-R)/2$ audio data encoded based on G711U-LAW Standard is transmitted through this channel.

A description 706 indicates a capability for receiving audio G.723.1 compressed based on G.723.1 Standard being one of the audio signal compression systems. These descriptions are described together with their parameters (not shown).

A description 707 indicates nonstandard audio data. Here, the $(L-R)/2$ audio data encoded based on G723.1 Standard is transmitted through this channel.

In the conventional video conference system only corresponding to monaural, the audio G.711A-LAW

09870910.033101
T076070980

(description 702), the audio G.711U-LAW (description 703) or the audio G.723.1 (description 706) may be selected in the capability selected. Namely, since the contents of the descriptions 704, 705 and 707 being the nonstandard audio are nonstandard, it is unnecessary to understand them, and any erroneous operation does not occur due to these descriptions.

In Fig. 7, T.120 DESCRIPTION in the description 701 indicates one of standards for describing the data conference capability, the environment to be used, and the like, and H.221 in the description 704 indicates one of video and audio multiplying standards based on H.320 Standard.

Another end point B 602 similarly performs call setting to the MC 604, and then performs capability exchange to other terminals according to H.245 Standard. It is assumed that, like the end point A 601, the video conference system at the end point B 602 has the stereo audio processing capability. Further, another end point C 603 similarly performs call setting to the MC 604, and then performs capability exchange to other terminals according to H.245 Standard.

The end point C 603 merely has a monaural audio processing capability, and thus its capability table is shown in Fig. 8. In Fig. 8, a description 801 indicates a data conference capability, a description 802 indicates a capability for receiving audio G.711A-

09870910.053101
T07E50" 07607860

LAW, a description 803 indicates a capability for receiving audio G.711U-LAW, and a description 804 indicates a capability for receiving audio G.723.1. These descriptions are described together with their parameters shown rightward. A description 805 indicates CAPABILITY DESCRIPTORS in which the entry numbers of the capability table are sequentially described from the ability to which it is intended to give priority.

10 In Fig. 6, the MC 604 integrates all participants' capability sets, and describes two entries in the communication mode table to be transmitted based on a communication mode command, such that the end points A 601 and B 602 select stereo G.711 and the end point C 15 603 selects monaural G.711. Then, the MC 604 transmits the table to the respective end points (as indicated by arrows 609, 610 and 611). One of the two entries is to manage the (L+R)/2 audio signal, i.e., the monaural audio signal, and the other thereof is to manage the 20 (L-R)/2 audio signal. Entries 1 and 2 which are described in the communication mode table are shown as blocks 622 and 623, respectively.

The entry 1 622 shows SESSION ID (= 1) representing a session, SESSION DESCRIPTION (= audio) 25 representing the content of the session, DATA TYPE (= G.711 monaural) representing a data type, MEDIA CHANNEL (= MCA1 605) representing a multicasting address for

00070910.053104

transmitting audio data, and MEDIA CONTROL CHANNEL (= MCA2 606) representing a multicasting address for transmitting audio control data.

The entry 2 623 shows SESSION ID (= 2)

5 representing a session, SESSION DESCRIPTION (= audio)
representing the content of the session, DATA TYPE (= nonstandard (L-R)/2) representing a data type, MEDIA CHANNEL (= MCA3 607) representing a multicasting address for transmitting audio data, and MEDIA CONTROL
10 CHANNEL (= MCA4 608) representing a multicasting address for transmitting audio control data.

After then, each participant's terminal turns on its own audio to start multicasting. The end point A 601 transmits the (L+R)/2 audio data to the MCA 1 605
15 as indicated by numeral 612, and the control data for the (L+R)/2 audio data to the MCA 2 606 as indicated by numeral 615. Further, the end point A 601 transmits the (L-R)/2 audio data to the MCA 3 607 as indicated by numeral 618, and the control data for the (L-R)/2 audio
20 data to the MCA 4 608 as indicated by numeral 620.

Similarly, the end point B 602 transmits the (L+R)/2 audio data to the MCA 1 605 as indicated by numeral 613, and the control data for the (L+R)/2 audio data to the MCA 2 606 as indicated by numeral 616.
25 Further, the end point B 602 transmits the (L-R)/2 audio data to the MCA 3 607 as indicated by numeral 619, and the control data for the (L-R)/2 audio data to

00870910-053101

the MCA 4 608 as indicated by numeral 621. Since the end point C 603 only has the monaural audio processing capability, the end point C 603 transmits the monaural audio data to the MCA 1 605 as indicated by numeral 5 614, and the control data for the monaural audio data to the MCA 2 606 as indicated by numeral 617.

It is assumed that each of the end points A 601 and B 602 has a decoding capability for two channels, and the end point C 603 has a decoding capability for 10 one channel. The end point A 601 receives the multicast $(L+R)/2$ and $(L-R)/2$ audio data, and performs the predetermined process of Fig. 4 for the received two-channel audio data by using the audio codec within the video conference system so as to reproduce the 15 stereo audio. Similarly, the end point B 602 receives the multicast $(L+R)/2$ and $(L-R)/2$ audio data, and performs the predetermined process for the received two-channel audio data by using the audio codec within the video conference system so as to reproduce the 20 stereo audio.

Since the end point C 603 has the decoding capability for one channel, the end point C 603 receives the audio data of the entry 1 (SESSION ID = 1), and performs a conventional predetermined process 25 for the received data so as to reproduce the monaural audio signal.

As described above, according to the present

09870910-053401
T07ES0-07604860

embodiment, even in the multipoint conference in which the terminals each having the stereo audio processing capability and the terminals each having the monaural audio processing capability mixedly participate, it is possible between the terminals each having the stereo audio processing capability to transmit and receive the stereo audio.

This means that the video conference system having the stereo audio processing capability can participate in the multipoint conference without lowering its stereo audio processing capability to adjust it to the capability of another terminal. Further, the terminal having the stereo audio processing capability is not required to simultaneously support the monaural audio and the stereo audio (e.g., to generate the monaural audio data in addition to the stereo audio data) for the terminal only having the monaural audio processing capability. For this reason, it is unnecessary to increase a processing capability at each terminal, and it is unnecessary to expand a bandwidth on the network more than necessity. In such a condition, it is possible to achieve the multipoint conference using the stereo audio and create a sound field with full presence.

Next, a method by which the terminal having the stereo audio processing capability notifies a communication partner's side that this terminal has the

05310105370909

stereo audio processing capability will be explained hereinafter. Fig. 9 shows an RTCP (real time control protocol) packet which is transmitted by the terminal having the stereo audio processing capability, in the
5 above structures such as the multipoint connection, the conference participant's terminal, and the like.

Concretely, Fig. 9 shows a sender report (SR) of the RTCP packet for issuing a control request from the reception side to the transmission side. This packet
10 includes a header, transmission side's information, a reception report block, and a source description (SDES). In the header, information such as a real time protocol (RTP) (= version 2), the packet (= RTCP SR), a payload type (= 200), a packet length (= 12), SSRC, and
15 the like is described. Further, the SR shows an NTP time stamp, an RTP time stamp, a transmission (sender's) packet count, and a transmission (sender's) octet count, as the transmission side's information. In the reception report block, information such as
20 SSRC, packet loss, an arrival interval jitter, and the like is described. Although the SDES can include some items, the first item should be an SDES header.

In the SDES header, a version and a payload type are described. In the next SDES item, a host name
25 (CNAME) which is necessary to the RTCP packet is described. In the next SDES item, private extensions (PRIV) which represents the video conference system's

09870910 053101

own capability and audio devices being used is described, whereby it is possible to notify the partner's terminal of such information.

For example, the end point A 601 uses stereo
5 microphones as the audio input device when the conference starts. At this time, the audio data output by the end point A 601 is stereo audio.

In the SDES of the RTCP packet corresponding to the stereo audio data, it is described that the audio
10 is transmitted with two channels. Since the end point B 602 participating in the conference has the stereo audio processing capability, the end point B 602 receives the two-channel data, i.e., the (L+R) and (L-R) data, transmitted by the end point A 601 and thus
15 reproduces the stereo audio.

During the conference, when the end point A 601 changes the audio input device from the stereo microphones to headsets, the end point A 601 transmits the monaural audio data through the channel which was
20 used to transmit the (L+R) data before the audio input device was changed. Besides, the end point A 601 stops the data transmission through the channel which was used to transmit the (L-R) data before the audio input device was changed. Further, it is described in the
25 SDES of the RTCP packet corresponding to the audio channel that the number of audio channels is "1", and this description is notified to the reception side.

09670910-053101

On the other hand, the end point B 602 receives the audio RTCP packet transmitted from the end point A 601 and thus detects that the audio of the end point A 601 was changed from the stereo audio to the monaural audio. Thus, the end point B 602 stops the data reception from the L-R channels used till then.

As described above, since the transmission side (the end point A 601) notifies the reception side (the end point B 602) of the number of audio channels, even if the number of audio channels on the transmission side is frequently changed, the number of channels on the reception side can be easily changed only by turning on/off the L-R channels. Thus, the processing capability and the band on the network can be efficiently used.

Further, in the SDES of the RTCP packet concerning the audio transmitted by the end point A 601, the information of the used audio input device is described in addition to the number of audio channels. The other end point participating in the conference receives the RTCP packet and reads the information of the audio input device of the end point A 601, whereby it is possible to notify the user of the audio input device used on the communication partner's side through an application. Thus, the user can know through a display whether the received audio is the monaural audio or the stereo audio.

00870010.053104

Since the end point B 602 receives the monaural audio, if the end point B 602 intends to request the stereo audio to the end point A 601, the end point B 602 sends a notification such that the end point A 601 transmits L-R data in response to a mode request of H.245 Standard. Thus, the end point A 601 actually generates and transmits the L-R audio data, whereby the end point B 602 can start the reception of the stereo audio.

As described above, the state that video conference system has the stereo audio processing capability is shown to the partner's terminal, whereby it is possible to easily and automatically change the number of audio channels during the conference.

According to the present embodiment, even in the multipoint conference in which the video conference and video telephone systems each having the stereo audio processing capability and the video conference and video telephone systems each having the monaural audio processing capability mixedly participate, it is possible between the video conference and video telephone systems each having the stereo audio processing capability to transmit and receive the stereo audio. This means that the video conference system having the stereo audio processing capability can participate in the multipoint conference without lowering its stereo audio processing capability to

adjust it to the capability of another system or terminal.

Further, the system or terminal having the stereo audio processing capability is not required to generate
5 the monaural audio data in addition to the stereo audio data for the system or terminal only having the monaural audio processing capability. For this reason, it is unnecessary to increase a processing capability at each system or terminal, and it is unnecessary to
10 expand a bandwidth on the network more than necessity. In such a condition, it is possible to efficiently use communication lines, achieve the multipoint conference using the stereo audio, and create a sound field with full presence.

15 Further, it is assumed that, in the communication between the video conference systems each having the stereo audio processing capability, the transmission side's terminal has the monaural audio input device and the stereo audio input device, and changes these two
20 kinds of audio input devices. In such a case, if one audio channel is changed to two audio channels, the transmission side's terminal notifies the communication partner of the information concerning the audio source change and the channel number change by using the PRIV
25 of the RTCP, and the reception side's terminal (the communication partner) turns on/off the L-R channels in response to the received notification, whereby it is

09870910 053101
"01604860"

possible to dynamically change the audio process between the terminals from the monaural audio process to the stereo audio process.

[Second Embodiment]

5 Next, Fig. 14 shows topology of the group telephone and conference according to the centralized multipoint connection. A communication system in the present embodiment is basically the same as that in the first embodiment, but in the present embodiment the MCU
10 has the specific feature corresponding to a stereo format.

 In Fig. 14, numeral 1501 denotes an MCU (multipoint control unit) corresponding to the stereo format in the present invention. The MCU 1501 has a
15 stereo signal processing capability and can perform the communication in the stereo communication system proposed in the first embodiment (this stereo communication system proposed in the first embodiment will be simply called the stereo communication system
20 hereinafter).

 In the stereo communication system, the $(L+R)/2$ signal (called a main audio signal hereinafter) obtained by the addition of the L and R audio signals and the $(L-R)/2$ signal (called a sub audio signal
25 hereinafter) obtained by the subtraction of the L and R audio signals are first encoded, and the stereo signal is managed by using the encoded data, whereby the

communication is performed.

In the data communication, the main audio signal is managed as the data being the G.723.1-encoded monaural audio to which the payload type has been
5 defined.

Since the sub audio signal can not be managed as the conventional audio data, the nonstandard payload type is allocated thereto in the audio data communication.

10 The MCU 1501 is composed of one MC (multipoint controller) and one MP (multipoint processor) for processing audio data.

Three terminals A 1502, B 1503 and C 1504 participate in the group telephone and conference, and
15 each terminal is point-point connected to the MCU 1501.

The terminals A 1502 and B 1503 are the video telephone and conference terminals corresponding to the stereo forma in the present invention. Like the MCU 1501, these terminals can perform the communication in
20 the previously proposed stereo communication system.

Since the terminal C 1504 is the conventional video telephone and conference terminal, the audio of this terminal is monaural.

First, a procedure to start the group telephone
25 and conference will be explained.

In order to start the group telephone and conference, the MC existing in the MCU 1501 performs

09870910-053101

setting to convene the conference.

09870910-053101
The terminal A 1502 performs call setting to the MC, and then performs capability exchange to other terminals according to H.245 Standard. Then, the
5 terminal A 1502 transmits a capability table as shown in Fig. 16 to the MC so as to show the MC that this terminal can perform the communication with the conventional audio processing capability (monaural audio processing capability) and the stereo
10 communication system.

The capability table in Fig. 16 will be briefly explained. A description 1701 indicates a data conference capability, a description 1702 indicates a capability for receiving audio G.711A-LAW, and a
15 description 1703 indicates a capability for receiving audio G.711U-LAW. The capabilities indicated by the descriptions 1702 and 1703 are the capability for transmitting monaural audio of one channel based on G.711 Standard. The terminal A 1502 transmits the main
20 audio signal by using this capability.

A description 1704 indicates a nonstandard audio data capability. Here, the sub audio signal encoded based on G711A-LAW Standard is managed. A description 1705 indicates a nonstandard audio data capability.
25 Here, the sub audio signal encoded based on G711U-LAW Standard is managed.

A description 1706 indicates a capability for

A description 1707 indicates a nonstandard audio data capability. Here, the sub audio signal encoded based on G723.1 Standard is managed.

The terminal B 1503 is the terminal which corresponds to the stereo communication system, as well as the terminal A 1502. Thus, the terminal B 1503 similarly performs call setting to the MC and then performs capability exchange to other terminals according to H.245 Standard.

The terminal C 1504 which is the conventional
terminal for managing the monaural audio performs call
25 setting to the MC and then performs capability exchange
to other terminals according to H.245 Standard. In the

capability exchange, by using the capability table, the terminal C 1504 shows the MC that this terminal is the terminal for managing the monaural audio.

As described above, between the MC and each of all the terminals participating in the group telephone and conference, the call setting and the subsequent capability exchange end. Thus, the MC integrates the capabilities of all the participants and determines the audio format used for the MCU 1501 to perform multicasting.

After the capability exchange between the MC and each terminal ended, setting of audio channel communication is performed. By using the previously determined data format (an encoding system, the number of channels, etc.) between each terminal and the MCU 1501, each terminal and the MCU 1501 mutually open RTP and RTCP channels and start data transmission.

Namely, the main audio channel, and the data channel (the RTP channel) and the data control channel (the RTCP channel) for the sub audio channel are respectively opened between the terminal using the stereo communication system and the MCU 1501.

On the other hand, only the data channel (the RTP channel) and the data control channel (the RTCP channel) for the main audio (monaural audio) are opened between the terminal managing the monaural signal and the MCU 1501, but any channel for the sub audio is not

opened (such the channel can not be opened due to the terminal's capability). Therefore, unnecessary data in the LAN can be prevented from increasing. However, for example, in a case where the data quantity does not increase, or in a case where all the terminals participating in the group telephone and conference communicate by using the stereo communication system, the main and sub audio data may be communicated through one channel.

Next, internal blocks of the terminal A 1502 will be briefly explained.

Fig. 11 shows the internal blocks in the terminal A 1502. Here, the terminal A 1502 is the video telephone and conference terminal which has the two audio channels for the L and R audio signals.

This terminal is controlled by a system controller 1205, and a video codec 1203 and an audio codec 1204 perform encoding and decoding of the respective data.

Programs for the system controller 1205, the video codec 1203 and the audio codec 1204 have been stored in a flash ROM 1207. Thus, after turning on a power supply, the system controller 1205 reads its program, loads it in an SDRAM 1208, and starts initialization of the terminal A 1502.

The programs for the video codec 1203 and the audio codec 1204 are read by the system controller 1205, loaded in an SRAM within each codec chip, whereby

the programs start.

The audio is input through stereo microphones, a line input, headsets, a wireless telephone connected by a wireless unit 1211, and the like.

5 Information selected by a user is input to the terminal through an USB (Universal Serial Bus) I/F 1206, an RS-232C (Recommended Standard 232C) I/F 1210 or a LAN I/F 1209; and based on the input information the system controller 1205 select the audio input
10 source by an audio input selector 1213.

The selected audio signal is digitized by an audio AD/DA converter 1212 and then input to the audio codec 1204.

For example, the audio codec 1204 performs audio
15 data compression based on G723.1 Standard.

The compressed audio data is sent to the system controller 1205, subjected to a predetermined process, and then sent to a LAN through the LAN I/F 1209.

On the other hand, in the data reception, the data
20 received through the LAN I/F 1209 is subjected to a predetermined process by the system controller 1205, and thus obtained audio data is sent to the audio codec 1204. If there is the video data, this data is sent to the video codec 1203.

25 The audio data is decoded by the audio codec 1204, converted into an analog signal by the audio AD/DA converter 1212, and output to an audio output device

00870910.053101
T0510.053101

Next, an internal audio data process in the video telephone and conference terminal (the terminal A 1502) will be explained with reference to Fig. 12.

5

10

15

20

25

The main audio data received from the MCU 1501 is decoded by a decoder 1304, and thus a main audio signal

1312 is output. The sub audio data received from the MCU 1501 is decoded by a decoder 1305, and thus a sub audio signal 1313 is output.

5 The main audio signal or the sub audio signal is the main audio signal or the sub audio signal obtained by appropriately compositing the audio of each of all the terminals A 1502, B 1503 and C 1504. Namely, the audio of the terminal A 1502 is composited in the main or sub audio signal. For this reason, it is necessary to reproduce the audio signal from which the audio of the terminal A 1502 has been eliminated, in order to prevent howling tones.

10 Thus, the main audio signal 1310 from the terminal A 1502 and the main audio signal 1312 from the MCU 1501 obtained by compositing the audio of each of all the terminals are input to an audio signal elimination block 1306, whereby the audio signal of the terminal A 1502 is eliminated.

15 An audio signal 1314 output from the audio signal elimination block 1306 is the signal obtained by compositing the audio signals of the terminals B 1503 and C 1504.

20 Also, since the audio signal 1314 is the monaural signal, it is possible to output this signal 1314 when the audio output of the terminal is monaural such as the headset or the like.

25 Similarly, the sub audio signal 1311 from the

09870940.05340.1

terminal A 1502 and the sub audio signal 1313 from the MCU 1501 are input to an audio signal elimination block 1307, whereby the sub audio signal of the terminal A 1502 is eliminated.

5 In the audio signal elimination block, the audio signal of its own terminal is eliminated by, e.g., an elimination method using correlation of the audio signals.

10 An output signal 1315 of the audio signal elimination block 1307 and the main audio signal 1314 are input to an arithmetic unit 1304. The arithmetic unit 1304 performs simple arithmetic operations for these signals and outputs the L and R audio signals.

15 Thus, when the audio output of the terminal A 1502 is stereo such as the speaker or the like, the L and R audio signals are output, whereby the stereo signal can be reproduced.

20 Next, Fig. 13 shows an audio data processing method in a monaural terminal such as the terminal C 1504.

25 The audio signal of the terminal is encoded by an encoder 1401, and then transmitted to the MCU 1501. Further, the received main audio data is decoded by a decoder 1402, and then input to an audio signal elimination block 1403 so as to eliminate the terminal's own audio. Thus, the signal from which the terminal's own audio has been eliminated is output from

09870910 053101
TOTALS 07602960

the audio signal elimination block 1403, and this audio signal is managed as the monaural audio output signal.

Next, the internal process of the MCU 1501 will be explained.

5 As shown in Fig. 15, the MCU 1501 receives the plural audio data from the three terminals. Namely, main and sub audio data 1505 are received from the terminal A 1502, main and sub audio data 1506 are received from the terminal B 1503, and monaural audio data 1507 is received from the terminal C 1504.

Fig. 10 shows the audio process within the MCU 1501.

15 The MCU 1501 decodes the plural received data, adds main and sub audio data to the decoded data, encodes the addition-result data, and performs multicasting of the encoded data.

20 Concretely, the following three kinds of audio signals, i.e., the main audio signal of the terminal A 1502 decoded by a decoder 1101, the main audio signal of the terminal B 1503 decoded by a decoder 1102, and the monaural signal of the terminal C 1504 decoded by a decoder 1103, are input to an adder 1106 which performs the addition of the main audio signals.

25 Further, the following two kinds of audio signals, i.e., the sub audio signal of the terminal A 1502 decoded by a decoder 1104, and the sub audio signal of the terminal B 1503 decoded by a decoder 1105, are

input to an adder 1107 which performs the addition of the sub audio signals.

A main audio signal 1508 output from the adder 1106 which performs the addition of the main audio signals is encoded by an encoder 1108, and then multicast from the MCU 1501 to the respective terminals. An example of a packet of the data multicast from the MCU 1501 is shown in Fig. 17.

The packet shown in Fig. 17 corresponds to one-channel monaural data of 8kHz sampling encoded according to G.711U-LAW Standard. Since the payload type of this data is defined as "0", values "0" are described in a payload type 1801 of the packet.

Further, a sub audio signal 1509 output from the adder 1107 which performs the addition of the sub audio signals is encoded by an encoder 1109, and then multicast from the MCU 1501 to the respective terminals. An example of a packet of the data multicast from the MCU 1501 is shown in Fig. 18.

The packet shown in Fig. 18 corresponds to one-channel audio data of 8kHz sampling encoded according to G.711U-LAW Standard. Since this data is obtained by encoding a difference signal between the L and R audio signals, only this data itself can not be reproduced as the audio signal. For this reason, this data is defined as the nonstandard audio, and the payload type of this data is dynamically allocated, i.e., values

"96" are described in a payload type 1901 of the packet in Fig. 18.

Each of the terminals A 1502 and B 1503 reproducing the stereo signal receives the multicast
5 main audio signal (Fig. 17) and sub audio signal (Fig. 18), and can reproduce the received signals as the stereo signal by using the blocks shown in Fig. 12.

On the other hand, the terminal C 1504 reproducing the monaural signal receives only the multicast main
10 audio signal (Fig. 17), and can reproduce the received audio of the group telephone and conference as the monaural signal by eliminating the terminal C's own audio.

As explained above, according to the present
15 embodiment, the MCU 1501 corresponding to the stereo format of the present invention performs the mutual communication of the audio data by using the stereo communication system. By doing so, even if the
20 terminals corresponding to the stereo signal and the terminals corresponding to the monaural signal are mixedly connected mutually, the terminal corresponding to the stereo signal can manage the stereo signal without matching its capability with the capability of the terminal corresponding to the monaural signal.
25 Besides, the terminal corresponding to the monaural signal can participate in the group telephone and conference using such the mutual communication, as it

09870910-053401
TOTESQ-07607860

remains its conventional function.

[Third Embodiment]]

As an MCU corresponding to a stereo format in the present embodiment, the function of the MCU in the
5 second embodiment is achieved by one of the terminals participating in the group telephone and conference.

Fig. 19 shows connection in a case where, when a stereo terminal A 11001, a stereo terminal B 11002 and a monaural terminal C 11003 together perform the group
10 telephone and conference, the terminal A 11001 achieves the MCU function within the terminal itself. In Fig. 19, the terminal A 11001 having the MCU function is point-point connected to the terminal B 11002, and the terminal A 11001 is further point-point connected to
15 the terminal C 11003.

The terminal A 11001 is the video telephone and conference terminal corresponding to the stereo format according to the present invention, and the terminal C 11003 is the conventional video telephone and
20 conference terminal of which audio is managed by the monaural audio signal.

A procedure to start the group telephone and conference is as follows.

In order to start the group telephone and
25 conference, an MC existing in the terminal A 11001 and being the part of the MCU function performs setting to convene the conference.

09870910-053101
T0E80 0T60960

09870910 "053101
T0562865

The terminal A 11001 performs call setting to the MC existing in the terminal A 11001 itself, and then performs capability exchange to other terminals according to H.245 Standard. Then, the terminal A

5 11001 transmits a capability table to the MC so as to show the MC that this terminal can perform the communication with the conventional audio processing capability (monaural audio processing capability) and the communication according to the stereo communication

10 system.

Next, the terminal B 11002 similarly performs call setting to the MC existing in the terminal A 11001, and then performs capability exchange to other terminals according to H.245 Standard. Then, the terminal B

15 11002 transmits a capability table to the MC so as to show the MC that this terminal can perform the communication with the conventional monaural audio processing capability and the communication according to the stereo communication system.

20 Next, the terminal C 11003 similarly performs call setting to the MC existing in the terminal A 11001, and then performs capability exchange to other terminals according to H.245 Standard. In the capability exchange, by using a capability table, the terminal C

25 11003 shows the MC that this terminal is the terminal for managing the monaural audio.

As described above, between the MC and each of all

the terminals participating in the group telephone and conference, the call setting and the subsequent capability exchange according to H.2445 Standard end. Thus, the MC integrates the capabilities of all the participants and determines the audio format used for the MCU (i.e., the terminal A 11001) to perform multicasting.

After the capability exchange between the MC and each terminal ended, setting of audio channel communication is performed. By using the previously determined data format (an encoding system, the number of channels, etc.) between each terminal and the MCU, the MCU and the terminal B 11002 and the MCU and the terminal C 11003 mutually open RTP and RTCP channels and start data transmission.

Namely, the main audio channel, and the data channel (the RTP channel) and the data control channel (the RTCP channel) for the sub audio channel are respectively opened between the terminal B 11002 using the stereo communication system and the MCU (the terminal A 11001). The data to be transmitted from the terminal B 11002 to the MCU (the terminal A 11001) is the main audio data and the sub audio data (11004), and the data to be transmitted from the terminal A 11001 to the terminal B 11002 is main audio data and sub audio data 11006 in which the participant's audio of the group telephone and conference is composited.

On the other hand, only the data channel (the RTP channel) and the data control channel (the RTCP channel) for the main audio (monaural audio) are opened between the terminal C 11003 managing the monaural signal and the MCU (the terminal A 11001), but any channel for the sub audio is not opened (such the channel can not be opened due to the terminal's capability). Thus, the data to be transmitted from the terminal C 11003 to the MCU (the terminal A 11001) is monaural data 11005, and the data to be transmitted from the terminal A 11001 to the terminal C 11003 is the main audio data (monaural data) 11007 in which the participant's audio of the group telephone and conference is composited.

Since the data transmitted from the terminal A 11001 to the terminal C 11003 may be only the main audio data, unnecessary data in the LAN can be prevented from increasing. However, for example, in a case where the data quantity does not increase, or in a case where all the terminals participating in the group telephone and conference communicate by using the stereo communication system, the main and sub audio data may be communicated through one channel.

Next, internal blocks of the terminal A 11001 will be briefly explained with reference to Fig. 20. As described above, the terminal A 11001 is the video telephone and conference terminal which corresponds to

00070910-053101

the stereo format and has the MCU function.

The terminal A 11001 is the terminal having the stereo signal processing capability. The L and R audio signals are input as the audio input, and the main and sub audio signals of this terminal itself are generated by an arithmetic unit 11101.

On the other hand, as the data received from another terminals, the main audio signal is input from the terminal B 11002, and the monaural audio data is received from the terminal C 11003. The main audio signal input from the terminal B 11002 is decoded by a decoder 11102 and input to an adder 11105, and the monaural audio data input from the terminal C 11003 is decoded by a decoder 11103 and input to the identical adder 11105. The audio of the terminal B 11002 and the audio of the terminal C 11003 are composited and thus the composited audio signal is output by the adder 11105. This audio signal is also the monaural signal which is output as the audio from the terminal A 11001.

As the sub audio signal received from another terminal, the sub audio signal received from the terminal B 11002 is decoded by a decoder 11104 and input to an adder 11106. Since there is no other input to the adder 11106, the sub audio signal of the terminal B 11002 is output as it is. This output signal from the adder 11106 is also the sub audio signal which is output as the audio from the terminal A

09870910 053104

The audio output signal of the terminal A 11001 is generated from the output signal of the adder 11105 obtained by compositing the main audio signal of the terminal B 11002 and the monaural signal of the terminal C 11003, and the output signal of the adder 11106 being the sub audio signal of the terminal B 11002. The audio signals output from the adders 11105 and 11106 are input to an arithmetic unit 11111, whereby the L and R audio output signals for stereo reproduction can be obtained from the main and sub audio signals. Since the terminal A 11001 has the MCU function, as described above, any block for eliminating the audio signal of this terminal itself is unnecessary, whereby it is possible to remarkably reduce a quantity of the operations.

Namely, in order to composite the main audio
20 signal of the terminal A 11001 to the output signal of
the adder 11105, the above two signals are input to an
adder 11107. The output from the adder 11107 is
encoded by an encoder 11109 according to a
predetermined encoding method, whereby the main audio
25 data to be broadcast can be obtained. On the other
hand, the output from the adder 11106 and the sub audio
signal of the terminal A 11001 are input to an adder

11108 for audio compositing. The output from the adder 11108 is encoded by an encoder 11110 according to a predetermined encoding method, whereby the sub audio data to be broadcast can be obtained. In the present embodiment, the main audio data is transmitted to the terminal B 11002 and the terminal C 11003, and the sub audio data is transmitted to only the terminal B 11002.

The terminal B 11002 receives the audio-composited main and sub audio data from the terminal A 11001, decodes the received data, eliminates the audio of the terminal B itself, and restores the L and R audio signals, whereby the stereo signal can be reproduced.

On the other hand, the terminal C 11003 receives only the audio-composited main audio data from the terminal A 11001, decodes the received data, eliminates the audio of the terminal C itself, and restores the audio, whereby the monaural signal can be reproduced.

As described above, even in the group telephone and conference in which the stereo and monaural terminals mixedly participate, the stereo terminal can perform the communication of the stereo audio, and the conventional monaural terminal can perform the communication of the monaural audio without providing any additional function.

The present invention includes a situation that a memory medium storing program codes of software to realize the functions of the above embodiments is

05370910-053101
T07E50" 07602950

of the L and R channels constituting the stereo audio
and the data obtained by subtraction of these audio
signals. Thus, in the multipoint conference in which
the terminal devices each having the stereo audio
5 processing capability and the terminal devices each
having the monaural audio processing capability mixedly
participate, it is possible between the terminal
devices each having the stereo audio processing
capability to restore or reproduce the stereo audio
10 without increasing a data quantity and wastefully
increasing processing capabilities.

Further, according to the present invention, even
if the video telephone and conference terminal
corresponding to the stereo format which performs by
15 using the stereo communication system the communication
of the data (main audio data) obtained by the addition
of the two L and R audio signals and the data (sub
audio data) obtained by the subtraction of these two
audio signals and the conventional terminal which has
20 the monaural signal processing capability mixedly
exist, it is possible to perform the communication of
the stereo format.

Further, even if the terminals managing the stereo
signal and the terminals managing the monaural signal
25 are mixedly connected mutually, the MCU of the present
invention necessary in the group telephone and
conference can manage the stereo signal without

09870910-053101

matching its capability with the capability of the terminal corresponding to the monaural signal (i.e., without integrating stereo audio and monaural audio into only monaural audio).

- 5 Further, the terminal performing the monaural signal process can participate in the group telephone and conference, as it remains its conventional function.

- 10 The present invention is not limited to the above embodiments. Namely, it is obvious that various modifications and changes are possible in the present invention within the spirit and scope of the appended claims.

09870910.053401
T01E50" DT60960